

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: BASF Aktiengesellschaft
- (B) STREET: Carl-Bosch-Strasse 38
- (C) CITY: Ludwigshafen
- (D) STATE: Rheinland Palatinate
- (E) COUNTRY: Federal Republic of Germany
- (F) POSTAL CODE: D-67056

(ii) TITLE OF APPLICATION: A process for preparing chiral carboxylic acids from nitriles using a nitrilase or microorganisms which comprise a gene for the nitrilase

(iii) NUMBER OF SEQUENCES: 9

(iv) COMPUTER-READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1071 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Double
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) [sic] ANTISENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Alcaligenes faecalis*
- (B) STRAIN: 1650

(ix) FEATURES:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1071

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG CAG ACA AGA AAA ATC GTC CGG GCA GCC GCC GTA CAG GCC GCC TCT
Met Gln Thr Arg Lys Ile Val Arg Ala Ala Ala Val Gln Ala Ala Ser
1 5 10 15

0050/49462-01000

29

CCC	AAC	TAC	GAT	CTG	GCA	ACG	GGT	GTT	GAT	AAA	ACC	ATT	GAG	CTG	GCT	96
Pro	Asn	Tyr	Asp	Leu	Ala	Thr	Gly	Val	Asp	Lys	Thr	Ile	Glu	Leu	Ala	
			20					25					30			
CGT	CAG	GCC	CGC	GAT	GAG	GGC	TGT	GAC	CTG	ATC	GTG	TTT	GGT	GAA	ACC	144
Arg	Gln	Ala	Arg	Asp	Glu	Gly	Cys	Asp	Leu	Ile	Val	Phe	Gly	Glu	Thr	
		35					40					45				
TGG	CTG	CCC	GGA	TAT	CCC	TTC	CAC	GTC	TGG	CTG	GGC	GCA	CCG	GCC	TGG	192
Trp	Leu	Pro	Gly	Tyr	Pro	Phe	His	Val	Trp	Leu	Gly	Ala	Pro	Ala	Trp	
		50					55					60				
TCG	CTG	AAA	TAC	AGT	GCC	CGC	TAC	TAT	GCC	AAC	TCG	CTC	TCG	CTG	GAC	240
Ser	Leu	Lys	Tyr	Ser	Ala	Arg	Tyr	Tyr	Ala	Asn	Ser	Leu	Ser	Leu	Asp	
		65				70				75					80	
AGT	GCA	GAG	TTT	CAA	CGC	ATT	GCC	CAG	GCC	GCA	CGG	ACC	TTG	GGT	ATT	288
Ser	Ala	Glu	Phe	Gln	Arg	Ile	Ala	Gln	Ala	Ala	Arg	Thr	Leu	Gly	Ile	
				85				90						95		
TTC	ATC	GCA	CTG	GGT	TAT	AGC	GAG	CGC	AGC	GGC	GGC	AGC	CTT	TAC	CTG	336
Phe	Ile	Ala	Leu	Gly	Tyr	Ser	Glu	Arg	Ser	Gly	Gly	Ser	Leu	Tyr	Leu	
			100					105					110			
GGC	CAA	TGC	CTG	ATC	GAC	GAC	AAG	GGC	GAG	ATG	CTG	TGG	TCG	CGT	CGC	384
Gly	Gln	Cys	Leu	Ile	Asp	Asp	Lys	Gly	Glu	Met	Leu	Trp	Ser	Arg	Arg	
		115					120					125				
AAA	CTC	AAA	CCC	ACG	CAT	GTA	GAG	CGC	ACC	GTA	TTT	GGT	GAA	GGT	TAT	432
Lys	Leu	Lys	Pro	Thr	His	Val	Glu	Arg	Thr	Val	Phe	Gly	Glu	Gly	Tyr	
		130					135				140					
GCC	CGT	GAT	CTG	ATT	GTG	TCC	GAC	ACA	GAA	CTG	GGA	CGC	GTC	GGT	GCT	480
Ala	Arg	Asp	Leu	Ile	Val	Ser	Asp	Thr	Glu	Leu	Gly	Arg	Val	Gly	Ala	
					150					155					160	
CTA	TGC	TGC	TGG	GAG	CAT	TTG	TCG	CCC	TTG	AGC	AAG	TAC	GCG	CTG	TAC	528
Leu	Cys	Cys	Trp	Glu	His	Leu	Ser	Pro	Leu	Ser	Lys	Tyr	Ala	Leu	Tyr	
				165					170					175		
TCC	CAG	CAT	GAA	GCC	ATT	CAC	ATT	GCT	GCC	TGG	CCG	TCG	TTT	TCG	CTA	576
Ser	Gln	His	Glu	Ala	Ile	His	Ile	Ala	Ala	Trp	Pro	Ser	Phe	Ser	Leu	
			180					185					190			
TAC	AGC	GAA	CAG	GCC	CAC	GCC	CTC	AGT	GCC	AAG	GTG	AAC	ATG	GCT	GCC	624
Tyr	Ser	Glu	Gln	Ala	His	Ala	Leu	Ser	Ala	Lys	Val	Asn	Met	Ala	Ala	
		195					200					205				
TCG	CAA	ATC	TAT	TCG	GTT	GAA	GGC	CAG	TGC	TTT	ACC	ATC	GCC	GCC	AGC	672
Ser	Gln	Ile	Tyr	Ser	Val	Glu	Gly	Gln	Cys	Phe	Thr	Ile	Ala	Ala	Ser	
		210					215				220					

105040-9280861

30

AGT GTG GTC ACC CAA GAG ACG CTA GAC ATG CTG GAA GTG GGT GAA CAC 720
 Ser Val Val Thr Gln Glu Thr Leu Asp Met Leu Glu Val Gly Glu His
 225 230 235 240
 AAC GCC CCC TTG CTG AAA GTG GGC GGC GGC AGT TCC ATG ATT TTT GCG 768
 Asn Ala Pro Leu Leu Lys Val Gly Gly Gly Ser Ser Met Ile Phe Ala
 245 250 255
 CCG GAC GGA CGC ACA CTG GCT CCC TAC CTG CCT CAC GAT GCC GAG GGC 816
 Pro Asp Gly Arg Thr Leu Ala Pro Tyr Leu Pro His Asp Ala Glu Gly
 260 265 270
 TTG ATC ATT GCC GAT CTG AAT ATG GAG GAG ATT GCC TTC GCC AAA GCG 864
 Leu Ile Ile Ala Asp Leu Asn Met Glu Glu Ile Ala Phe Ala Lys Ala
 275 280 285
 ATC AAT GAC CCC GTA GGC CAC TAT TCC AAA CCC GAG GCC ACC CGT CTG 912
 Ile Asn Asp Pro Val Gly His Tyr Ser Lys Pro Glu Ala Thr Arg Leu
 290 295 300
 GTG CTG GAC TTG GGG CAC CGA GAC CCC ATG ACT CGG GTG CAC TCC AAA 960
 Val Leu Asp Leu Gly His Arg Asp Pro Met Thr Arg Val His Ser Lys
 305 310 315 320
 AGC GTG ACC AGG GAA GAG GCT CCC GAG CAA GGT GTG CAA AGC AAG ATT 1008
 Ser Val Thr Arg Glu Glu Ala Pro Glu Gln Gly Val Gln Ser Lys Ile
 325 330 335
 GCC TCA GTC GCT ATC AGC CAT CCA CAG GAC TCG GAC ACA CTG CTA GTG 1056
 Ala Ser Val Ala Ile Ser His Pro Gln Asp Ser Asp Thr Leu Leu Val
 340 345 350
 CAA GAG CCG TCT TGA 1071
 Gln Glu Pro Ser
 355

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Thr Arg Lys Ile Val Arg Ala Ala Ala Val Gln Ala Ala Ser
 1 5 10 15
 Pro Asn Tyr Asp Leu Ala Thr Gly Val Asp Lys Thr Ile Glu Leu Ala
 20 25 30

Arg	Gln	Ala	Arg	Asp	Glu	Gly	Cys	Asp	Leu	Ile	Val	Phe	Gly	Glu	Thr	
		35					40						45			
Trp	Leu	Pro	Gly	Tyr	Pro	Phe	His	Val	Trp	Leu	Gly	Ala	Pro	Ala	Trp	
	50					55					60					
Ser	Leu	Lys	Tyr	Ser	Ala	Arg	Tyr	Tyr	Ala	Asn	Ser	Leu	Ser	Leu	Asp	
65					70					75					80	
Ser	Ala	Glu	Phe	Gln	Arg	Ile	Ala	Gln	Ala	Ala	Arg	Thr	Leu	Gly	Ile	
				85					90					95		
Phe	Ile	Ala	Leu	Gly	Tyr	Ser	Glu	Arg	Ser	Gly	Gly	Ser	Leu	Tyr	Leu	
			100					105					110			
Gly	Gln	Cys	Leu	Ile	Asp	Asp	Lys	Gly	Glu	Met	Leu	Trp	Ser	Arg	Arg	
		115					120					125				
Lys	Leu	Lys	Pro	Thr	His	Val	Glu	Arg	Thr	Val	Phe	Gly	Glu	Gly	Tyr	
	130					135					140					
Ala	Arg	Asp	Leu	Ile	Val	Ser	Asp	Thr	Glu	Leu	Gly	Arg	Val	Gly	Ala	
145					150					155					160	
Leu	Cys	Cys	Trp	Glu	His	Leu	Ser	Pro	Leu	Ser	Lys	Tyr	Ala	Leu	Tyr	
				165					170					175		
Ser	Gln	His	Glu	Ala	Ile	His	Ile	Ala	Ala	Trp	Pro	Ser	Phe	Ser	Leu	
			180					185					190			
Tyr	Ser	Glu	Gln	Ala	His	Ala	Leu	Ser	Ala	Lys	Val	Asn	Met	Ala	Ala	
		195					200					205				
Ser	Gln	Ile	Tyr	Ser	Val	Glu	Gly	Gln	Cys	Phe	Thr	Ile	Ala	Ala	Ser	
	210					215					220					
Ser	Val	Val	Thr	Gln	Glu	Thr	Leu	Asp	Met	Leu	Glu	Val	Gly	Glu	His	
225					230					235					240	
Asn	Ala	Pro	Leu	Leu	Lys	Val	Gly	Gly	Gly	Ser	Ser	Met	Ile	Phe	Ala	
				245					250					255		
Pro	Asp	Gly	Arg	Thr	Leu	Ala	Pro	Tyr	Leu	Pro	His	Asp	Ala	Glu	Gly	
			260					265					270			
Leu	Ile	Ile	Ala	Asp	Leu	Asn	Met	Glu	Glu	Ile	Ala	Phe	Ala	Lys	Ala	
		275					280					285				
Ile	Asn	Asp	Pro	Val	Gly	His	Tyr	Ser	Lys	Pro	Glu	Ala	Thr	Arg	Leu	
	290					295					300					
Val	Leu	Asp	Leu	Gly	His	Arg	Asp	Pro	Met	Thr	Arg	Val	His	Ser	Lys	
305					310					315					320	

32

Ser Val Thr Arg Glu Glu Ala Pro Glu Gln Gly Val Gln Ser Lys Ile
 325 330 335

Ala Ser Val Ala Ile Ser His Pro Gln Asp Ser Asp Thr Leu Leu Val
 340 345 350

Gln Glu Pro Ser
 355

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Peptide

(iii) HYPOTHETIC: NO

(iii) [sic] ANTISENSE: NO

(v) FRAGMENT TYPE: N Terminus

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Alcaligenes faecalis*
- (B) STRAIN: 1650

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Nitrilase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Gln Thr Arg Lys Ile Val Arg Ala Ala Ala Val Gln Ala Ala Ser
 1 5 10 15

Pro Asn Tyr Asp Leu Ala Thr Gly Val Asp Lys Thr Ile Glu Leu Ala
 20 25 30

Arg Gln Ala Arg Asp Glu Gly
 35

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Peptide

(iii) HYPOTHETIC: NO

(iii) [sic] ANTISENSE: NO

0050/49462-040504

33

(v) FRAGMENT TYPE: Internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Alcaligenes faecalis*

(B) STRAIN: 1650

(vii) IMMEDIATE SOURCE:

(B) CLONE: Nitrilase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Glu	Glu	Ala	Pro	Glu	Gln	Gly	Val	Gln	Ser	Lys	Ile	Ala	Ser	Val	Ala
1				5				10						15	

Ile	Ser	His	Pro	Gln
				20

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 Amino acids

(B) TYPE: Amino acid

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Peptide

(iii) HYPOTHETICAL: NO

(iii) [sic] ANTISENSE: NO

(v) FRAGMENT TYPE: Internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Alcaligenes faecalis*

(B) STRAIN: 1650

(vii) IMMEDIATE SOURCE:

(B) CLONE: Nitrilase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Glu	Glu	Ala	Pro	Glu	Gln	Gly	Val	Gln	Ser	Lys
1				5				10		

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 Base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

005049462-92890860

34

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) [sic] ANTISENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Alcaligenes faecalis*
 - (B) STRAIN: 1650
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: Nitrilase
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATGCAGACNA GNAARATCGT SCG

23

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 Base pairs
 - (B) TYPE: Nucleic acid
 - (C) STRANDEDNESS: Single
 - (D) TOPOLOGY: Linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) [sic] ANTISENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Alcaligenes faecalis*
 - (B) STRAIN: 1650
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: Nitrilase
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TNGCSACNGA NGCRATCTTG

20

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 Base pairs
 - (B) TYPE: Nucleic acid
 - (C) STRANDEDNESS: Single
 - (D) TOPOLOGY: Linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) [sic] ANTISENSE: NO

35

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Alcaligenes faecalis*
- (B) STRAIN: 1650

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Nitrilase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTAATCATAT GCAGACAAGA AAAATCGTCC G

31

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) [sic] ANTISENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Alcaligenes faecalis*
- (B) STRAIN: 1650

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Nitrilase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AAGGATCCTC AAGACGGCTC TTGCACTAGC AG

32

0050/49462